intensity values (MM_i) for the transcript, where each of the PM_i is paired with one of the MM_i; calculating a p-value using one sided Wilcoxon's signed rank test, where the p-value is for a null hypothesis that $\theta=$a threshold value and an alternative hypothesis that said $\theta>$ the threshold value, wherein said θ is a test statistic for intensity difference between the perfect match intensity values and mismatch intensity values; and indicating whether the transcript is present based upon the p-value.

Please replace the paragraph on page 5, lines 12-15 with the following:

In some particularly preferred embodiments, the testing statistic is $median((PM_i-MM_i)/(PM_i+MM_i))$. In these embodiments, the threshold value is a constant. Typically, the threshold value is around 0.001 to 0.05. Most preferably, the threshold value is around 0.015.

Please replace the paragraph on page 6, lines 8-18 with the following:

The presence, marginal presence or absence (detected, marginally detected or undetected) of a transcript may be called based upon the p–value and significance levels. Significance levels, α_1 and α_2 may be set such that: $0<\alpha_1<\alpha_2<0.5$. Note that for the one-sided test, if null hypothesis is true, the most likely observed p-value is 0.5, which is equivalent to 1 for the two-sided test. Let p be the p-value of one-sided signed rank test. In preferred embodiments, if $p<\alpha_1$, a "detected" call can be made (i.e., the expression of the target gene is detected in the sample). If $\alpha_1 \leq p <\alpha_2$, a marginally detected call may be made. If $p\geq\alpha_2$, "undetected call" may be made. The proper choice of significance levels and the thresholds can reduce false calls. In some preferred embodiments,

$0<\alpha_1<\alpha_2<0.06$. In some particularly preferred embodiments, α_1 is around 0.04 and α_2 is around 0.06.

Please replace the paragraph on page 7, lines 11-16 with the following:

In some particularly preferred embodiments of the computer software products of the invention, the testing statistic is *median((PM_i-MM_i)/(PM_i+MM_i))* and threshold value is a constant. The computer program product may contain code for accepting user's selection or input of the threshold value. A default value may be used as well. Typically, the threshold value is around 0.001 to 0.05. In a particularly preferred embodiment, the threshold value is around 0.015.

Please replace the paragraph on page 7, lines 17-22 with the following:

The presence, marginal presence or absence (detected, marginally detected or undetected) of a transcript may be called based upon the p –value and significance levels. Significance levels, α_1 and α_2 may be set such that: $0<\alpha_1<\alpha_2<0.5$. In preferred embodiments, if $p<\alpha_1$, a "detected" call can be made (i.e., the expression of the target gene is detected in the sample). If $\alpha_1 \le p <\alpha_2$, a marginally detected call may be made. If $p \ge \alpha_2$, "undetected call" may be made. The proper choice of significance levels and the

Please replace the paragraph on page 8, lines 3-10 with the following:

The computer software product may include computer program code for indicating that the transcript is present, absent or marginally absent. The computer program code, when executed, may indicate the result by causing the display of the result

on a display device such as a screen. Alternatively, the result may be outputted into a file. In addition, the result may be temporarily stored in a computer memory device so that other computer program modules may access this result. In some preferred embodiments, the computer software products may include code to accept a user's selection of various significance levels.

The computer software product may include computer program code for indicating that the transcript is present, absent or marginally absent. The computer program code, when executed, may indicate the result by causing the display of the result on a display device such as a screen. Alternatively, the result may be outputted into a file. In addition, the result may be temporarily stored in a computer memory device so that other computer program modules may access this result. In some preferred embodiments, the computer software products may include code to accept a user's selection of various significance levels.

In addition, systems for determining whether a transcript is present in a biological sample are also provided. The systems include a processor; and a memory being coupled to the processor, the memory storing a plurality of machine instructions that cause the processor to perform a plurality of logical steps when implemented by the processor; the logical steps include the method steps of the invention.

α⁶

Methods for making and using molecular probe arrays, particularly nucleic acid

probe arrays are also disclosed in, for example, U.S. Patent Numbers 5,143,854,

5,242,974, 5,252,743, 5,324,633, 5,384,261, 5,405,783, 5,409,810, 5,412,087, 5,424,186,

5,429,807, 5,445,934, 5,451,683, 5,482,867, 5,489,678, 5,491,074, 5,510,270, 5,527,681,

5,527,681, 5,541,061, 5,550,215, 5,554,501, 5,556,752, 5,556,961, 5,571,639, 5,583,211,

5,593,839, 5,599,695, 5,607,832, 5,624,711, 5,677,195, 5,744,101, 5,744,305, 5,753,788,

5,770,456, 5,770,722, 5,831,070, 5,856,101, 5,885,837, 5,889,165, 5,919,523, 5,922,591,

5,925,517, 5,658,734, 6,022,963, 6,150,147, 6,147,205, 6,153,743, 6,140,044 and

D430024, all of which are incorporated by reference in their entireties for all purposes.

Typically, a nucleic acid sample is labeled with a signal moiety, such as a fluorescent

label. The sample is hybridized with the array under appropriate conditions. The arrays

are washed or otherwise processed to remove non-hybridized sample nucleic acids. The

hybridization is then evaluated by detecting the distribution of the label on the chip. The

distribution of label may be detected by scanning the arrays to determine florescence

intensities distribution. Typically, the hybridization of each probe is reflected by several

pixel intensities. The raw intensity data may be stored in a gray scale pixel intensity file.

The GATC™ Consortium has specified several file formats for storing array intensity

data. The final software specification is available at the Consortium's website and is

incorporated herein by reference in its entirety. The pixel intensity files are usually large.

For example, a GATC™ compatible image file may be approximately 50 Mb if there are

about 5000 pixels on each of the horizontal and vertical axes and if a two byte integer is

used for every pixel intensity. The pixels may be grouped into cells (see, GATC™

Please replace the paragraph on page 17, lines 9-15 with the following:

The embodiments of the invention will be described using GeneChip® high oligonucleotide density probe arrays (available from Affymetrix, Inc., Santa Clara, CA, USA) as exemplary embodiments. One of skill in the art would appreciate that the embodiments of the invention are not limited to high density oligonucleotide probe arrays. In contrast, the embodiments of the invention are useful for analyzing any parallel large scale biological analysis, such as those using nucleic acid probe arrays, protein arrays, etc.

Please replace the paragraph on page 18, lines 1-7 with the following:

in several patents previously incorporated by reference. In such embodiments, a single square-shaped feature on an array contains one type of probe. Probes are selected to be specific against desired target. Methods for selecting probe sequences are disclosed in, for example, U.S. Patent Application Nos. 09/718,295, 09/721,042, and 60/252,617, all incorporated herein by reference in their entireties for all purposes.

Please replace the paragraph on page 20, lines 14-21 with the following:

Computer software products may be written in any of various suitable programming languages, such as C, C++, C# (Microsoft®), Fortran, Perl, MatLab (MathWorks), SAS, SPSS and Java. The computer software product may be an independent application with data input and data display modules. Alternatively, the computer software products may be classes that may be instantiated as distributed objects. The computer software products may also be component software such as Java

Beans (Sun Microsystems), Enterprise Java Beans (EJB, Sun Microsystems), Microsoft®

COM/DCOM (Microsoft®), etc.

Please replace the paragraph on page 23, lines 5-11 with the following:

In some embodiments, Wilcoxon's signed rank test is used to analyze paired PM and MM probes. In a block of n probe pairs (also known as atoms, Figure 3) for detecting a gene (typically 10, 15, or 20 probe pairs). Each probe pair typically consists of two cells, one has the sequence designed to be perfectly matching the target sequence and the other has the sequence designed to be mismatching the target sequence, preferably at only a single nucleotide location (usually at the center of the sequence segment).

Please replace the paragraph on page 23, lines 12-20 with the following:

Let the i-th perfectly matching cell intensity be PM_i and the i-th mismatching cell intensity be MM_i (i=1,....,n). All these data are positive numbers. As described above, in some embodiments, the hybridization of each probe may be reflected by several pixel intensities. In such embodiments, the cell intensity is derived from the pixel intensities. In preferred embodiments, around 60, 70, 75, 80, 85, or 90 percentile of intensities of inner pixels in a cell is used to represent the cell intensity. In a particularly preferred embodiment, the 75 percentile of intensities of inner pixels in a cell is used to represent the cell intensity and is saved in a CEL file together with the number of pixels and the standard deviation of intensities at these pixels.

Please replace the paragraph on page 29, lines 3-12 with the following:

a^{11}

In some particularly preferred embodiments, the following three statistics of cell intensities can be used to make calls based on one sided Wilcoxon's signed rank test. The null hypothesis is denoted H_0 and alternative hypothesis H_1.

(1) H_0: median $(PM_i - MM_i) = \tau_1$;

        H_1: median $(PM_i - MM_i) > \tau_1$

(2) H_0: median $(PM_i - MM_i)/ (PM_i + MM_i) = \tau_2$;

        H_1: median $(PM_i - MM_i)/ (PM_i + MM_i) > \tau_2$;

(3) H_0: median $(PM_i - B_i) = \tau_3$;

        H_1: median $(PM_i - B_i) > \tau_3$;

Please replace the paragraph on page 32, lines 15-22 with the following:

a^{12}

The presence, marginal presence or absence (detected, marginally detected or undetected) of a transcript may be called based upon the p–value and significance levels (54-58). Significance levels, α_1 and α_2 may be set such that: $0<\alpha_1<\alpha_2<0.5$. Note that for the one-sided test, if null hypothesis is true, the most likely observed p-value is 0.5, which is equivalent to 1 for the two-sided test. Let p be the p-value of one sided signed rank test. In preferred embodiments, if $p<\alpha_1$, a "detected" call can be made (i.e., the expression of the target gene is detected in the sample). If $\alpha_1 \leq p <\alpha_2$, a marginally detected call may be made. If $p \geq \alpha_2$, "undetected call" may be made. The proper choice of significance